

What is claimed is:

1. A method for assembling nucleic acid base sequences comprising the steps of:

moving a window of fixed length along a first nucleic acid base sequence and simultaneously searching for a second nucleic acid base sequence which has a partial sequence at a terminal region thereof matching with a sequence defined by the window;

determining whether the second nucleic acid base sequence searched in said step and the first nucleic acid base sequence can be assembled or not; and

assembling said first nucleic acid base sequence and said second nucleic acid bases sequence if said step determines that the second nucleic acid base sequence and the first nucleic acid base sequence can be assembled.

2. A method for assembling nucleic acid base sequences comprising the steps of:

moving a window of fixed length along a first nucleic acid base sequence and simultaneously searching for a second nucleic acid base sequence which has a partial sequence at a terminal region thereof matching with a sequence defined by the window;

determining whether the second nucleic acid base sequence searched in said step and the first nucleic acid base sequence can be assembled or not; and

assembling said first nucleic acid base sequence and said second nucleic acid base sequence if said step determines that the second nucleic acid base sequence and the first nucleic acid base sequence can be assembled,

SEARCHED
INDEXED
COPIED
FILED

wherein the nucleic acid base sequence assembled in said step is used as a new first nucleic acid base sequence to repeatedly carry out said steps.

3. A method for assembling nucleic acid sequences comprising the steps of:

entering identification information about each of a plurality of nucleic acid base sequences and a fixed-length partial sequence located in a terminal region of the nucleic acid base sequence into a table, both of which are associated with each other;

constructing a first consensus sequence based on a first nucleic acid base sequence;

searching for a nucleic acid base sequence which has a partial sequence matching with a part of said consensus sequence with reference to said table;

comparing a sequence adjacent to said partial sequence of the nucleic acid base sequence searched in said step with a sequence adjacent to said partial sequence of said consensus sequence, and determining whether the searched nucleic acid base sequence can be assembled to said consensus sequence or not; and

assembling said nucleic acid base sequence to said consensus sequence so as to reconstruct a consensus sequence if said step determines that the nucleic acid base sequence can be assembled to the consensus sequence.

4. The method for assembling nucleic acid base sequences according to claim 3, wherein a sequence whose base length is the longest among unprocessed nucleic acid base sequences is selected as said first nucleic acid base sequence.

5. A method for assembling nucleic acid base sequences comprising:

a first step of sorting a plurality of nucleic acid base sequences in descending order of their sequence lengths;

a second step of entering identification information about each of the plurality of nucleic acid base sequences and a fixed-length partial sequences located in the head end and tail end regions of input nucleic acid base sequences into a table, both of which are associated with each other;

a third step of selecting one of the nucleic acid base sequences whose sequence length is the longest among the plurality of unprocessed nucleic acid base sequences, and constructing a first consensus sequence;

a fourth step of moving a fixed length window along said consensus sequence and simultaneously searching for an unprocessed nucleic acid base sequence which has a partial sequence matching with a sequence defined by said fixed length window with reference to said table;

a fifth step of comparing said consensus sequence with the unprocessed nucleic acid base sequence searched in said fourth step, and determining whether the both sequences can be assembled or not; and

a sixth step of assembling the nucleic acid base sequence searched in said fourth step to said consensus sequence so as to reconstruct a consensus sequence if said fifth step determines that the both sequences can be assembled,

wherein the fourth step to the sixth step are repeated until said fixed length window completes the scanning throughout said

consensus sequence, and said third step to said sixth step are repeated if any unprocessed nucleic acid base sequence still remains.

6. The method for assembling nucleic acid base sequences according to claim 3 comprising a step of specifying a number of said fixed base length of partial sequences to be entered into said table for one nucleic acid base sequence.

7. The method for assembling nucleic acid base sequences according to claim 3 comprising a step of designating a range of the head end and tail end regions of said nucleic acid base sequence from which said fixed-length partial sequences to be entered into said table is extracted.

8. The method for assembling nucleic acid base sequences according to claim 3 wherein a base length of said fixed base length of partial sequence to be entered into said table is at least 10 bases or more and 32 bases or less.

9. The method for assembling nucleic acid base sequences according to claim 3 comprising:

a step of specifying an upper limit c on the expected value of the number of entries which are detected upon once making reference to said table and their associated input sequences are determined not to be able to be assembled to said consensus sequences, and

a step of specifying the length s of fixed-length partial sequences to be entered into said table as an integer satisfying the following expression (1)

$$s \geq \frac{1}{2} \log \frac{KN}{c} \quad \dots(1)$$

where N is the number of said plurality of input nucleic acid base sequences and K is the number of the fixed-length partial sequences selected from each nucleic acid base sequence.

10. The method for assembling nucleic acid base sequences according to claim 3 wherein two-way lists are used to implement said consensus sequences.

11. The method for assembling nucleic acid base sequences according to claim 3 wherein a said fixed length of partial sequence is represented by a fixed number of computing words which are independent of the length of the fixed-length partial sequences.

12. The method for assembling nucleic acid base sequences according to claim 3 wherein only entries corresponding to a key which occurs a previously specified number of times or less in said table are utilized.